WOODHOUSE EXHIBIT 2

EXHIBIT A

Message
_____

From: Eleonora Presani [█████████@meta.com]
Sent: 10/19/2022 4:09:56 PM
To: Eleonora Presani [█████████@meta.com]; Joelle Pineau [█████████@meta.com]
Subject: Message summary [{"otherUserFbId":100021673387183,"threadFbId":null}]

Joelle Pineau (10/19/2022 05:26:01 PDT):
>Hi! Very sorry I haven't gotten back to you yet on this. It's on my to-do list, and will definitely
not drop it!

Eleonora Presani (10/19/2022 05:26:40 PDT):
>Hi! Sorry I didn't update you, we managed to find a poc.

Eleonora Presani (10/19/2022 05:26:53 PDT):
>Mary Williamson helped me

Joelle Pineau (10/19/2022 05:26:53 PDT):
>That's awesome!!! Who?

Eleonora Presani (10/19/2022 05:27:06 PDT):
>One researcher per area

Eleonora Presani (10/19/2022 05:27:19 PDT):
>For the SRT submission, for this round of work...

Eleonora Presani (10/19/2022 05:27:46 PDT):
>We still need your help to find someone who is going to help in the next rounds, like someone who can
own the process

Joelle Pineau (10/19/2022 05:29:06 PDT):
>Ok! Who are the researchers doing it this round?

Eleonora Presani (10/19/2022 05:31:55 PDT):
>Gabriel for Speech, will be covered by Jade while he's away

Eleonora Presani (10/19/2022 05:32:30 PDT):
>Aaron for CV, Luke for NLP and Randy Schultheis for embodied AI

Eleonora Presani (10/19/2022 05:34:16 PDT):
>For the ownership of the project going forward we were thinking TPMs would be the best fit, since they
have experience with working with XFN and collect dataset. Brian is in principle in agreement, but he
says he doesn't have people

Joelle Pineau (10/19/2022 05:36:50 PDT):
>Good to hear! That's very aligned with where I was going to push. Ideally TPMs, otherwise area leads
who are managers.

Joelle Pineau (10/19/2022 05:39:03 PDT):
>Slight change of topic: We're getting in some very complex territories with dataset use for our biggest
initiatives Zetta and M2C2 and I was wondering if you had some bandwidth to help, or if you're totally
full with other projects. This would be closely aligned with our top priority research effort in multi-
modal models, which sort of straddles research areas.

Eleonora Presani (10/19/2022 05:39:44 PDT):
>Would you have time to chat about it to tell me a bit more?

Eleonora Presani (10/19/2022 05:40:25 PDT):
>I've been working with Melanie for Zetta, for now I've been focusing on building an evaluation and
mitigations framework and I'm starting to think through a release strategy

Eleonora Presani (10/19/2022 05:41:41 PDT):
>So it's not too much of a stretch to help a bit with dataset, I've been talking about data with Melanie
as well because it's connected with the release strategy

Joelle Pineau (10/19/2022 05:44:02 PDT):

                               Redacted

                               Redacted   Melanie is driving
some of this already for Zetta, and Mary for M2C2, but I think your recent experience with responsible
strategy and PDTs would be really helpful.

Eleonora Presani (10/19/2022 05:44:40 PDT):
>I just heard about it yesterday from Melanie, I'm happy to take more of a closer look

Eleonora Presani (10/19/2022 05:45:56 PDT):
>Just to know where you stand, can I ask you this. Talking with Melanie we realized that probably one can only have 2 out of 3 among: openness, compliance and performances

Joelle Pineau (10/19/2022 05:48:29 PDT):
>That's exactly the discussion we need to have! There is an inherent tension between each of these. I think we don't have a choice on compliance, but sometimes risk and compliance are mistaken. We have flexibility on openness and performance. I'd also like to experiment with different ways to address openness, that respect compliance concerns but don't compromise research integrity; we do this routinely with health-care research, i.e. we don't release the dataset but we describe key characteristics.

Eleonora Presani (10/19/2022 05:50:07 PDT):
>I think that if we use third party data and then we don't open it up because we couldn't really use these data, that's a bit different

Eleonora Presani (10/19/2022 05:50:37 PDT):
>There may be reasons why we want to do it, but I am not totally comfortable doing it just because our competitors are doing it

Eleonora Presani (10/19/2022 07:02:29 PDT):
>related question: I've been helping with buying content from ███████ so through that I got in the conversation for Theorem Proving - and Zetta and other teams also need this type of content. One thing that came up is that they've been looking at dataset such as LibGen and SciHub. Now, I understand that coming from publishing I am biased, but I feel that using pirated material should be beyond our ethical threshold… what do you think?

Eleonora Presani (10/19/2022 08:58:50 PDT):
>btw, CONGRATS! on becoming the FAIR Lead!

Joelle Pineau (10/19/2022 09:03:40 PDT):
>Thx!

Joelle Pineau (10/19/2022 09:05:18 PDT):
>Yes, that's part of the conversation. For now we have agreed to use it just for pure exploration to see if there is value. If we find value, then we will setup proper licensing agreement. You think it's problematic to use even for this phase? I must say I don't have the full context on that material, so happy to understand better. | Redacted |
| Redacted |

Eleonora Presani (10/19/2022 09:09:56 PDT):
>I think it's ok to use it for exploratory purposes, but we need to be very careful to use only licensed data for the actual training of the model. SciHub, ResearchGate, LibGen are basically like PirateBay or something like that, they are distributing content that is protected by copyright and they're infringing it.